The Journal of Infectious Diseases Journal of Infectious Diseases Advance Access published January 18, 2016

MAJOR ARTICLE

Transcriptomic Analysis of the Host Response and Innate Resilience to Enterotoxigenic *Escherichia coli* Infection in Humans

William E. Yang,[1,2] Sunil Suchindran,[2] Bradly P. Nicholson,[2,3] Micah T. McClain,[2,3,4] Thomas Burke,[2] Geoffrey S. Ginsburg,[2] Clayton D. Harro,[5] Subhra Chakraborty,[5] David A. Sack,[5] Christopher W. Woods,[2,3,4] and Ephraim L. Tsalik[2,4,6]

[1]Duke University School of Medicine, [2]Center for Applied Genomics & Precision Medicine, Department of Medicine, Duke University, [3]Internal Medicine Service, Durham VA Medical Center, and [4]Division of Infectious Diseases, Department of Medicine, Duke University Medical Center, North Carolina; [5]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; and [6]Emergency Medicine Service, Durham VA Medical Center, North Carolina

Background. Enterotoxigenic *Escherichia coli* (ETEC) is a globally prevalent cause of diarrhea. Though usually self-limited, it can be severe and debilitating. Little is known about the host transcriptional response to infection. We report the first gene expression analysis of the human host response to experimental challenge with ETEC.

Methods. We challenged 30 healthy adults with an unattenuated ETEC strain, and collected serial blood samples shortly after inoculation and daily for 8 days. We performed gene expression analysis on whole peripheral blood RNA samples from subjects in whom severe symptoms developed (n = 6) and a subset of those who remained asymptomatic (n = 6) despite shedding.

Results. Compared with baseline, symptomatic subjects demonstrated significantly different expression of 406 genes highlighting increased immune response and decreased protein synthesis. Compared with asymptomatic subjects, symptomatic subjects differentially expressed 254 genes primarily associated with immune response. This comparison also revealed 29 genes differentially expressed between groups at baseline, suggesting innate resilience to infection. Drug repositioning analysis identified several drug classes with potential utility in augmenting immune response or mitigating symptoms.

Conclusions. There are statistically significant and biologically plausible differences in host gene expression induced by ETEC infection. Differential baseline expression of some genes may indicate resilience to infection.

Keywords. *E. coli*; immune response; gene expression; microarrays; bacterial infections; diarrheal illness.

Enterotoxigenic *Escherichia coli* (ETEC) is a common cause of diarrhea worldwide, especially in children and travelers. However, it is difficult to clinically distinguish from other causes of diarrhea, and diagnostic testing is limited in ETEC-prevalent areas. The majority of infections are self-limited though secondary dehydration can be lethal. Despite its prevalence, little is known about the host response to ETEC infection. Knowledge of the mechanisms behind clinical variability could assist in diagnosis, treatment, and prevention, because no clinically effective vaccine has been developed.

One approach to understand host-pathogen interactions is through a human challenge model. Although animal models facilitate the study of human disease, the ETEC strains that infect humans are not pathogenic in animals [1]. Consequently, human challenge studies have been the primary mechanism of studying human ETEC disease, which have revealed the dose dependence of symptoms, clinical variability between strains, and clinical course of infection [1, 2]. ETEC challenge studies have also been used to evaluate vaccine candidates and other treatments. Other pathogen challenge experiments have generated insights into the host response to infection. For example, human challenge with respiratory viruses have characterized gene expression changes that led to the development of a diagnostic signature [3].

Based on these prior successes, we performed a human ETEC challenge to study the transcriptomic response. Not all subjects became sick, allowing a comparison of symptomatic subjects to their own baseline and to asymptomatic subjects. Moreover, differences in baseline gene expression may reveal host factors conferring resilience to infection in asymptomatic individuals. This study design also facilitates a drug repositioning analysis whereby known drug-induced gene expression changes can be matched to the ETEC-induced response, identifying potential therapeutic targets.

METHODS

Challenge Protocol

Healthy subjects (age 23–45) were enrolled except in the case of likely exposure to ETEC or cholera within the past 2 years. The protocol was approved by all relevant institutional review boards and was in compliance with all applicable institutional and federal regulations governing human subjects' protection.

Received 28 July 2015; accepted 27 November 2015.
Correspondence: E. L. Tsalik, DUMC 102359, Durham, NC 27710 (e.t@duke.edu).

The Journal of Infectious Diseases®

DOI: 10.1093/infdis/jiv593

Subjects were admitted to the Johns Hopkins Bayview Medical Center, Center for Immunization Research inpatient unit. A stool sample was screened for infection at admission. Subjects were orally administered ETEC strain H10407 (O78:H11; colonization factor antigen [CFA]/I, labile toxin [LT]$^+$stable toxin [ST]$^+$) in a bicarbonate solution, with random assignment to a dose of either 1×10^5 or 1×10^6 colony-forming units, lower than described elsewhere [4]. This previously used strain is considered more virulent than other strains in the challenge model [1]. Blood was collected in PAXgene Blood RNA tubes (PreAnalytiX) early after ETEC challenge (8 hours), defining the baseline for this analysis. Subjects were monitored for approximately 8 days. Blood was collected, and symptoms were assessed daily. Ciprofloxacin was initiated on day 5 or sooner for severe symptoms.

Case Definitions

Patient-reported symptoms and stool output were recorded daily as described elsewhere [4]. Diarrhea was defined as 1 loose or liquid stool >300 g (grade 3 or greater) or >2 loose or liquid stools >200 g in any 48-hour period within 120 hours of ETEC challenge. Moderate-to-severe diarrhea was defined as ≥4 (or >400 g total) grade 3–5 stools in a 24-hour period.

Microarray Analysis

Total RNA was extracted from human blood using the PAXgene Blood miRNA Kit (Qiagen), according to the manufacturer's protocol. RNA quantity and quality were assessed using the Nanodrop spectrophotometer (ThermoScientific) and Agilent 2100 Bioanalyzer (Agilent), respectively. Hybridization and data collection were performed at Expression using the GeneChip Human Genome U133A2.0 Array (Affymetrix), according to the Affymetrix technical manual. Microarray data were Robust Multi-array Average normalized.

Statistical Methods

Bayesian factor regression modeling [5] was used to model correlation structure in the expression data. This served as a dimension-reduction step in which 20 000 gene expression measurements were expressed as 78 factors, each of which summarizes the expression of correlated genes into a single score. Each factor contains a variable number of genes, and any gene may be present in multiple factors. This analytic strategy has been used effectively in other challenge and gene expression studies [3]. The method is unsupervised (ie, does not use phenotype labels). A linear mixed model was used to compare symptomatic profiles and asymptomatic profiles over time. An analysis-of-covariance model, accounting for each subject's baseline gene expression, was used to study differences at the time of peak symptoms. A Benjamini–Hochberg false discovery rate (FDR) <10% accounted for multiple comparisons. Analysis was conducted using R 3.1 software (R Foundation for Statistical Computing, Vienna, Austria; available at: http://www.r-project.org/) [6, 7].

The *limma* method [8] was used to detect genes whose baseline difference persisted throughout the time course. A mixed model accounted for correlations within subjects and the *limma* empirical-Bayes method used information across genes for variance estimation. Genes with an FDR <5% were considered significant.

Functional Annotation

We used BioMart software (available at: http://www.biomart.org/) [9] to map probe set identifiers to HGNC gene symbols using the *Homo sapiens* GRCh38.p2 data set. DAVID 6.7 software (National Institutes of Health, Bethesda, MD; available at: http://david.ncifcrf.gov/) [10] was used for functional annotation. The gene ontology (GO) data sets queried included biological processes, cellular components, and molecular functions. Significance was defined at an FDR of <10%.

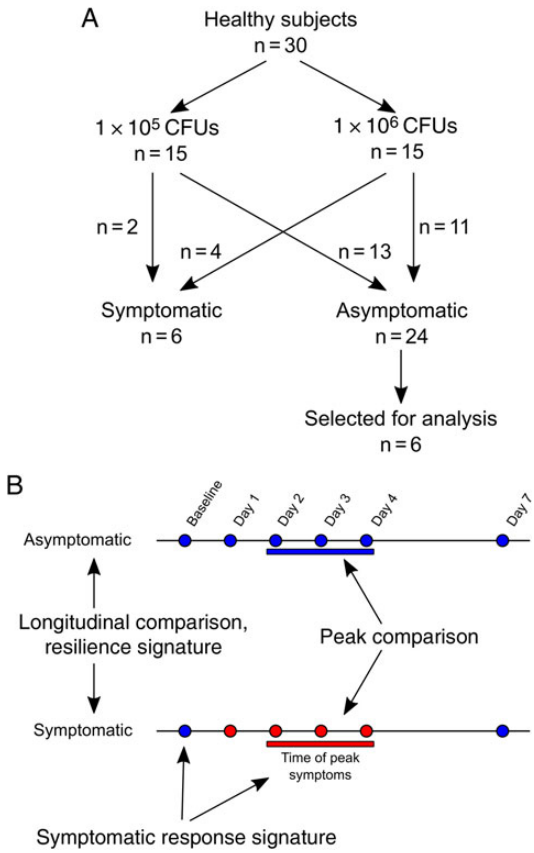

Figure 1. Experimental design. *A*, Thirty healthy subjects were randomly assigned to an enterotoxigenic *Escherichia coli* dose of either 1×10^5 or 1×10^6 colony-forming units (CFUs). Of the 15 subjects assigned to receive 1×10^5 CFUs, 2 became symptomatic and 13 remained asymptomatic. Of the 15 assigned to receive 1×10^6 CFUs, 4 became symptomatic and 11 remained asymptomatic. In all, 6 subjects had severe symptoms, and 24 remained asymptomatic. For analyses comparing symptomatic and asymptomatic subjects, only 6 asymptomatic subjects were included. *B*, The symptomatic response signature compared symptomatic subjects at the time of peak symptoms (which varied from day 2 to day 4) to their baseline (8 hours after challenge). The peak comparison compared symptomatic subjects at the time of peak symptoms with matched time points in asymptomatic subjects. The longitudinal comparison and resilience signature were derived from comparing all time points in symptomatic subjects with all time points in asymptomatic subjects.

RESULTS

Research Subjects

Subjects were randomly assigned to 2 challenge doses. Two of 15 subjects had severe symptoms at 1×10^5 colony-forming units, and 4 of 15 subjects at 1×10^6 colony-forming units (Figure 1*A*). All 6 subjects (3 female, 3 male) with severe symptoms underwent gene expression analysis, along with 6 asymptomatic subjects (1 female, 5 male). We matched for sex where possible, but only 1 asymptomatic female was available as a control. Peak symptoms occurred 2–4 days after challenge, and all 12 subjects demonstrated shedding of ETEC in stool at some point during the study (Figure 2; Supplementary Figure 1). Before challenge, subjects were screened for preexisting antibodies including lipopolysaccharide immunoglobulin (Ig) A and IgG, heat-labile toxin B (LTB) IgA and IgG, and CFA/I IgA and IgG. There were no significant differences in titer between symptomatic and asymptomatic subjects except for LTB IgG, for which asymptomatic subjects had higher levels ($P = .04$; 95% confidence interval, −502.9 to −9.849 ELISA units). However, these differences were unlikely to have affected symptom status, given the lack of protection conferred by such antibodies [11, 12].

Host Response in Symptomatic Infection

To characterize the host response to ETEC infection, we compared the gene expression of symptomatic subjects (n = 6) at peak symptoms compared with baseline (symptomatic response signature). Of the factors identified, 19 of 78 achieved significance at 10% FDR. We then defined the 1551 probe sets represented within these 19 factors. Although 19 factors achieved significance, not all 1551 probe sets did. Probe sets are grouped into factors because of their correlation with each other and not necessarily with the ETEC response. Thus, each probe set was analyzed for individual significance. Only 403 of 1551 probe sets achieved significance when comparing peak symptom to baseline gene expression (Supplementary Table 1).

These 403 probe sets mapped to 406 unique genes. Of the 403 probe sets, 152 were up-regulated and 251 were down-regulated at the time of peak symptoms. Functional annotation of the 152 up-regulated probe sets identified 7 GO categories (10% FDR) including vacuole biology and immune response (Table 1). Functional annotation of the 251 down-regulated probe sets identified 3 GO categories (10% FDR) (Table 1).

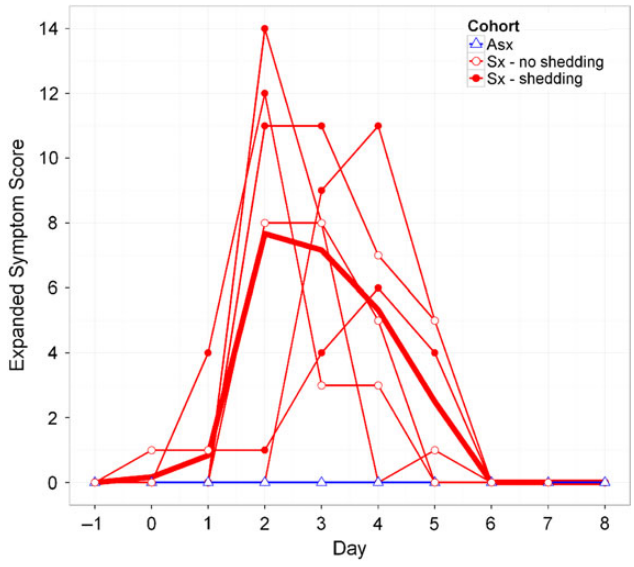

Figure 2. Symptom scores and shedding status over time for each individual subject. The thick line indicates average symptom score across all symptomatic (Sx) subjects within the cohort. Symptomatic subjects experienced peak symptoms between days 2 and 4. Asymptomatic (Asx) subjects maintained an expanded symptom score of 0 throughout the study. All subjects demonstrated shedding at some point during the study. For symptomatic subjects, solid symbols indicate detectable shedding of enterotoxigenic *Escherichia coli* in stool and empty symbols indicate undetectable shedding.

Symptomatic Versus Asymptomatic Comparison

Whereas all subjects were challenged with ETEC and shed the bacteria during the experiment, symptoms developed in only 6 of 12. To characterize the changes in gene expression between these 2 groups, 2 comparisons were used (Figure 1*B*). The first approach compared symptomatic subjects at peak symptoms to a corresponding time point in asymptomatic subjects (peak comparison), adjusted for each subject's own baseline gene expression. This comparison captures the changes in gene expression associated with symptomatic disease. In the peak comparison, 9 factors (822 probe sets) achieved significance. Of those 822 probe sets, 160 (118 up-regulated and 42 down-regulated) achieved individual significance ($P < .05$; Supplementary Table 2), mapping to 156 genes.

The second approach compared symptomatic and asymptomatic subjects over the entire 7-day study (longitudinal comparison). This approach captured differences that may have been present at times other than peak symptoms (Figure 3 offers an example). Fourteen factors (1014 probe sets) achieved significance. Of these 1014 probe sets, 137 achieved individual significance ($P < .05$; Supplementary Table 3), mapping to 135 genes. Of these 14 factors, 5 were also significant in the peak comparison (factors 1, 14, 29, 44, and 76). Factor 76 was most significant in both peak and longitudinal comparisons (Figure 4). Of the genes significant in factor 76, *ICAM1* is related to the class I major histocompatibility complex (MHC) pathway [13], and *ADM* encodes a protein with antibacterial activity against *E. coli* [14].

Combining the results from both comparisons maximized our ability to discriminate gene expression differences while maintaining statistical rigor. This resulted in 265 probe sets mapping to 254 unique genes. Functional annotation of the 265 probe sets resulted in 16 statistically significant GO categories, highlighting various aspects of the immune response (Table 1).

Resilience to Infection

To further characterize the difference between asymptomatic and symptomatic subjects, we hypothesized that host factors

Table 1. Statistically Significant Gene Ontology Categories[a]

Category	Term	Count	BH FDR
Up-regulated in the symptomatic response signature			
Biological process	Immune response	23	9.42×10^{-4}
	Regulation of myeloid cell differentiation	7	0.08396
	Leukocyte activation	11	0.091652
Cellular component	Lytic vacuole	14	3.84×10^{-5}
	Lysosome	14	3.84×10^{-5}
	Vacuole	14	1.42×10^{-4}
	Plasma membrane	50	0.075273
Down-regulated in the symptomatic response signature			
Biological process	Translation	17	0.036361
Cellular component	Ribonucleoprotein complex	21	0.03004
	Cytosolic part	11	0.028623
Symptomatic vs asymptomatic comparison			
Biological process	Immune response	42	6.34×10^{-10}
	Positive regulation of lymphocyte-mediated immunity	7	0.033853
	Positive regulation of leukocyte-mediated immunity	7	0.033853
	Regulation of lymphocyte-mediated immunity	8	0.038661
	Regulation of immune effector process	10	0.038934
	Defense response	25	0.035345
	Regulation of leukocyte-mediated immunity	8	0.035171
	Positive regulation of immune effector process	7	0.044825
	Leukocyte-mediated immunity	9	0.041318
	Positive regulation of immune response	11	0.048479
	Lymphocyte-mediated immunity	8	0.052921
	Positive regulation of immune system process	14	0.062121
	Immunoglobulin-mediated immune response	7	0.074069
	Adaptive immune response	8	0.07484
	Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	8	0.07484
	B-cell–mediated immunity	7	0.070551
Resilience signature (up-regulated in resilient cohort)			
Molecular function	MHC protein binding	4	1.53×10^{-5}
	MHC class I protein binding	3	0.001133

Abbreviations: BH FDR, Benjamini-Hochberg false discovery rate; MHC, major histocompatibility complex.

[a] Categories listed were statistically significant at 10% FDR. Counts represent the number of DAVID identifiers (mapped from probes) in each gene ontology category.

confer resilience to symptomatic infection. To test this hypothesis, we identified gene expression differences between susceptible (symptomatic) and resilient (asymptomatic) subjects that were detectable at baseline. Sex was included as a covariate given the mismatch between groups. Twenty-six probe sets mapping to 29 genes were statistically significant (5% FDR). Half were up-regulated in resilient compared with susceptible subjects (Figure 5, Supplementary Table 4).

Even at a 5% FDR, these genes may represent spurious associations. We therefore evaluated their biological plausibility in conferring resilience to infection. VarElect [13] uses GeneCards databases to associate genes with phenotypes of interest. We used VarElect to find links between the resilience signature and the "*Escherichia coli*" and "infection" phenotypes. Seven genes were directly linked to the *E. coli* phenotype, whereas the majority of remaining genes were indirectly related. Fifteen genes were directly related to the infection phenotype, including 6 of 7 related to the *E. coli* phenotype. These 6 genes fell into 2 groups: *C4BPA* (down-regulated in resilience) and 5 tubulin genes (up-regulated in resilience). *C4BPA* inhibits the classic complement pathway, and tubulins are involved in *E. coli* infection pathogenesis. Other genes directly related to the infection phenotype included *GZMH* and *TAP2*. *WARS* was associated with the *E. coli* phenotype. Functional annotation of the 13 down-regulated probe sets in resilient subjects identified no statistically significant GO categories, whereas the 13 up-regulated probe sets were enriched for genes in the MHC protein binding and MHC class I protein binding pathways (Table 1).

Drug Repositioning Analysis

The Broad Connectivity Map database [15] (Connectivity Map build 02) describes gene expression changes induced in human cell lines by 6100 perturbagens (small bioactive molecules). We hypothesized that comparing ETEC-induced changes to

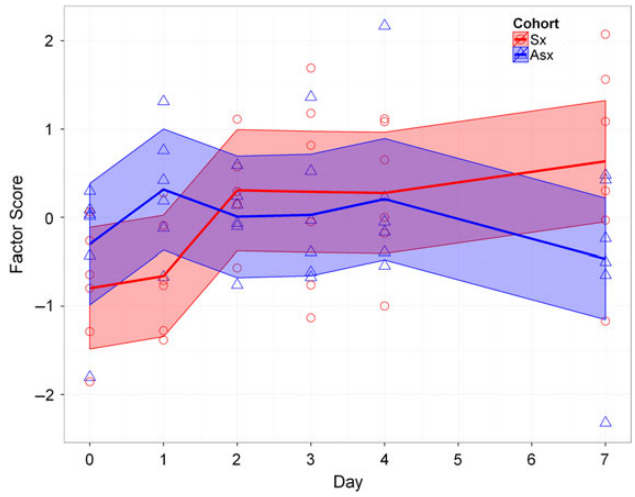

Figure 3. Longitudinal expression profile of factor 59. Shaded region indicates model mean and 95% confidence interval for each cohort. This is an example of a factor that was included in the longitudinal comparison ($P = .02$; adjusted for Benjamini–Hochberg false discovery rate [FDR]) but not the peak comparison ($P = .32$; FDR adjusted). The factor scores, representing gene expression, of both cohorts was comparable at the time of peak symptoms (varying by subject between days 2 and 4), and the difference in factor scores was greater at day 1 and day 7. Abbreviations: Asx, asymptomatic; Sx, symptomatic.

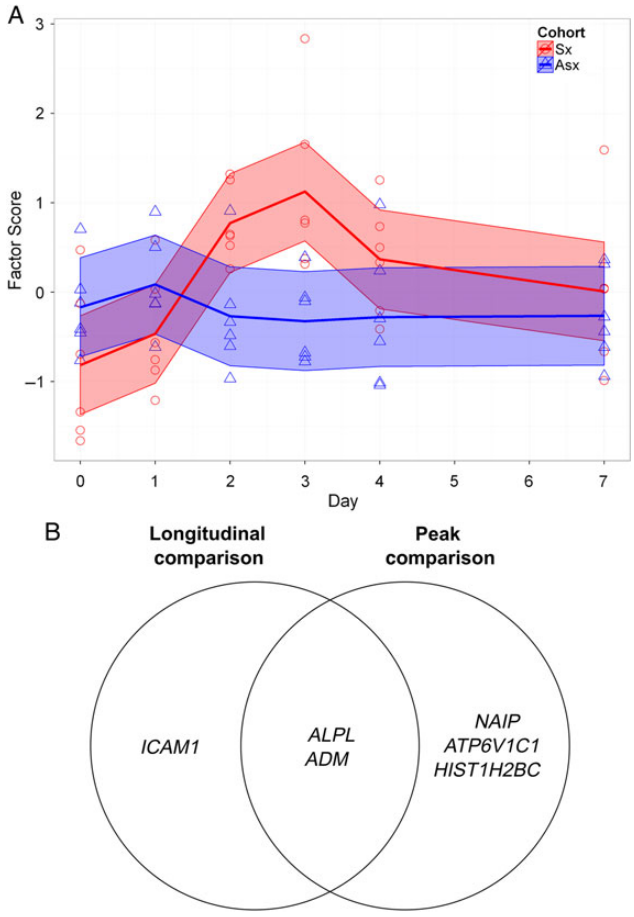

Figure 4. Factor 76 was the most significant factor present in both the longitudinal comparison ($P = .004$; adjusted for Benjamini–Hochberg false discovery rate [FDR])) and peak comparison ($P = .01$; FDR). *A*, Longitudinal expression profile of factor 76. Shaded region indicates model mean and 95% confidence interval for each cohort. *B*, Comparison of significant genes from factor 76. Although factor 76 was present in both comparisons, the component probe sets achieving individual statistical significance in each comparison were not identical. Probe sets were converted to genes using BioMart software. Abbreviations: Asx, asymptomatic; Sx, symptomatic.

perturbagen-induced changes could identify agents capable of mitigating or treating ETEC disease. Positive enrichment scores indicate positive correlation with ETEC-induced changes and could potentially augment the host response, mitigating disease. Negative enrichment scores indicate inverse correlations and may be used to reduce symptoms.

The results of drug repositioning analysis on the symptomatic response signature (97 perturbagens, Supplementary Table 5) and the peak comparison (110 perturbagens, Supplementary Table 6) were combined (Table 2) to identify perturbagens of interest. Among the most significant correlations were protein synthesis inhibitors (anisomycin and puromycin). Typical antipsychotics (thioridazine, fluphenazine, trifluoperazine, prochlorperazine, and perphenazine) and piperacetazine, an antipsychotic prodrug, were also significantly correlated with ETEC-induced changes. Zinc biology was represented by vorinostat and trichostatin A, histone deacetylase inhibitors that chelate zinc. 15-Δ-Prostaglandin J_2 [16] and withaferin A [17] inhibit the NF-κB pathway. Antibacterials were also identified, including cephalexin and amoxicillin with negative and rifampicin and metronidazole with positive enrichment scores.

DISCUSSION

Gene Expression Profiling

Although many details of ETEC pathogenesis are known, ours is the first study defining associated changes in peripheral blood gene expression. This analysis has identified gene expression differences at baseline that discriminate subjects in whom symptomatic ETEC infection develops from those who remain asymptomatic, revealing potential mechanisms of resilience to infection.

Many of the genes identified through our analyses have previously been identified in the host response to infection. *DHRS9, FCER1G, EMR1, CEACAM1,* and *CD177* were identified in a comparison of systemic inflammatory response syndrome and sepsis [18] though not always in the same direction of change as observed in this challenge. *MS4A4A, FEN1, SERPING1, OAS1, MCM6,* and *GINS2* were notably dysregulated in the host response to influenza A [3, 19]. *OAS1,* up-regulated in symptomatic subjects with ETEC, encodes an RNase that is part of the innate immune response to viral infection [14], raising the possibility of an immunoprotective role in bacterial or toxin-producing infections. *SERPING1,* also up-regulated in symptomatic subjects

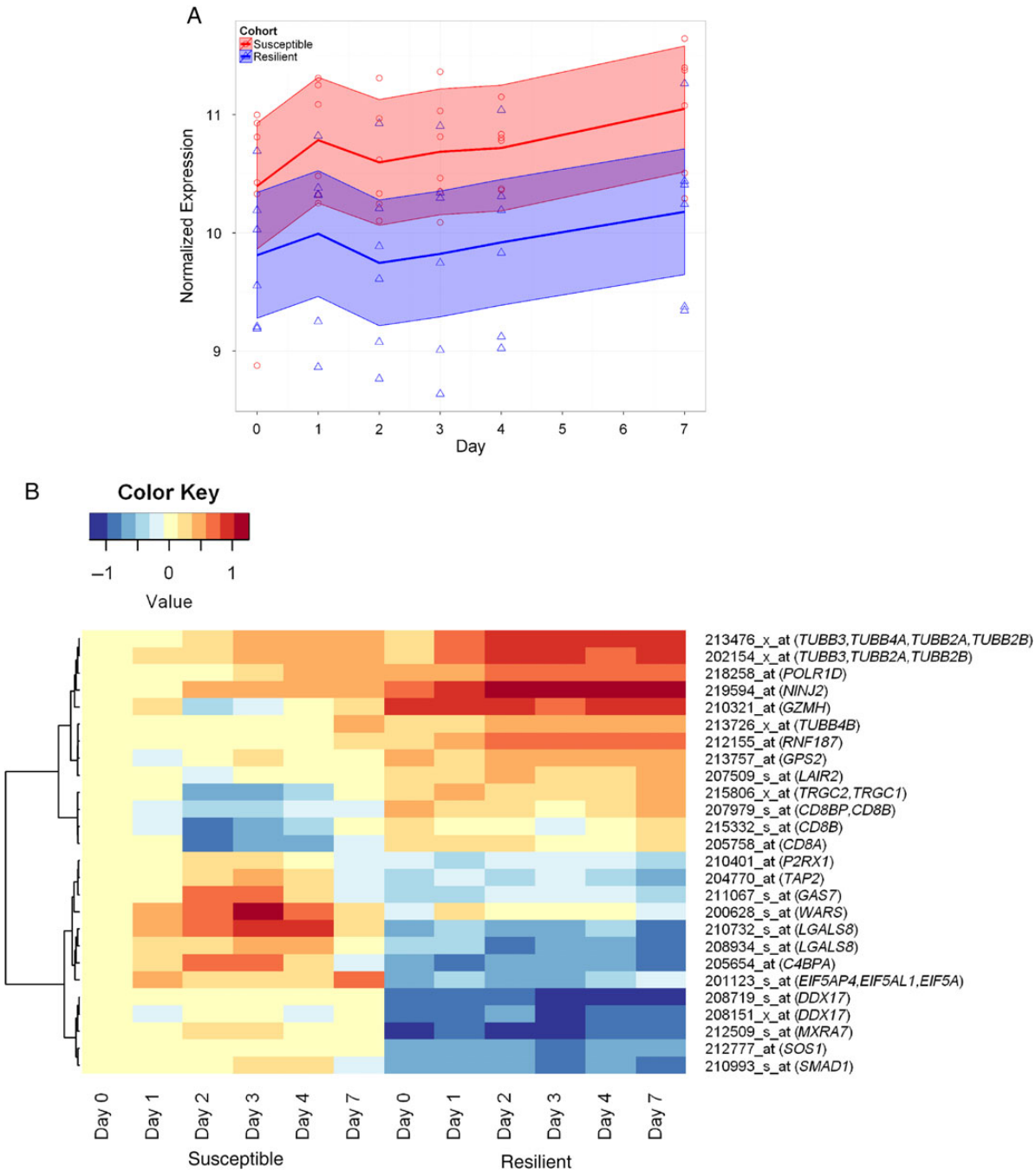

Figure 5. Resilience signature. *A, EIF5A* (201123_s_at) is an example of a gene with a constant difference in expression, suggesting resilience to enterotoxigenic *Escherichia coli* infection ($P = .03$; adjusted for Benjamini–Hochberg false discovery rate [FDR]). Shaded region indicates model mean and 95% confidence interval for each cohort. *B,* Heat map plot of all 26 probe sets identified in the signature, averaged across subjects. Values are relative to the susceptible day 0 average across subjects. Left half of the plot represents gene expression in susceptible subjects; right half, gene expression in resilient subjects. Probe sets were converted to genes using BioMart software.

with ETEC, inhibits the complement cascade [14]. Considering that ETEC is a toxin-mediated infection, the role of complement in ETEC is less obvious. However, C5a has proinflammatory properties, directly stimulating interleukin 6 production from colonic epithelial cells in a murine *Citrobacter rodentium* infection, which mimics human enteric *E. coli* infections [20].

Twenty-five genes from our analyses were also identified in a comparison of pediatric host response to invasive *E. coli* versus *Streptococcus pneumoniae* infection [19]. Many of these genes more strongly associated with the *S. pneumoniae* response than with nongastrointestinal *E. coli* infection. The reasons for these differences are unclear. On the surface, one might

Table 2. Drug Repositioning Analysis: Top 40 Statistically Most Significant Perturbagens, Ordered by Enrichment Score[a]

Perturbagen	Enrichment Score
Candidates to augment host response	
1,4-Chrysenequinone	0.978
Anisomycin	0.949
Albendazole	0.927
Prenylamine	0.897
Rottlerin	0.897
Anisomycin	0.859
Tonzonium bromide	0.826
Withaferin A	0.822
Helveticoside	0.821
Astemizole	0.817
Puromycin	0.817
H-7	0.814
5155877	0.813
Helveticoside	0.733
Monensin	0.703
Amitriptyline	0.7
Raloxifene	0.644
Vorinostat	0.635
Thioridazine	0.623
Trichostatin A	0.621
Wortmannin	0.59
Fluphenazine	0.568
15-Δ-Prostaglandin J_2	0.541
Trifluoperazine	0.518
Thioridazine	0.517
Prochlorperazine	0.489
Prochlorperazine	0.477
LY-294002	0.399
Tanespimycin	0.289
Candidates to mitigate pathogen-induced symptoms	
Pioglitazone	−0.602
PHA-00851261E	−0.623
Ethotoin	−0.698
Acacetin	−0.743
Amiodarone	−0.754
Procaine	−0.769
Scopolamine	−0.819
Piperine	−0.831
Fusaric acid	−0.837
Harmol	−0.878
7-Aminocephalosporanic acid	−0.882

[a] Scores range from −1 to +1. A positive enrichment score indicates that the gene expression pattern induced by the perturbagen in the database matches the direction of gene expression induced by enterotoxigenic *Escherichia coli* in the specified comparison. A negative enrichment score indicates an opposite direction. The magnitude of the score indicates the degree of similarity between the perturbagen-induced and enterotoxigenic *E. coli*–induced signatures. Some perturbagens may be listed more than once if they were significant in both the symptomatic response signature and the peak comparison.

assume that different *E. coli* pathovars induce similar host responses. However, both ETEC and *S. pneumoniae* cause disease in sites of frequent colonization: the intestine and respiratory tract, respectively. As such, the nature of their host-pathogen interactions are likely to be more similar than ETEC is to invasive *E. coli* infection.

This finding relates to one of the limitations in this study: We measured peripheral blood cell gene expression and not intestinal mucosal gene expression, which could not be obtained for safety reasons. Although we are unable to directly correlate gene expression in blood with intestinal mucosa, there are published data to suggest correlations are high, including in pigs, which are prone to postweaning ETEC-related gastroenteritis [21, 22]. In contrast to this human challenge, porcine challenges have largely focused on the mucosal response although some describe correlations between peripheral blood or lymphoid tissue and mucosal biology. For example, the increased number of $CD3^+CD4^+CD8^-$ T cells in the peripheral blood mirrored changes seen in small intestinal lamina propria [23]. Gene expression of many immune-related genes were equivalent in jejunal lamina propria and Peyer patches, from which immune cells are expected to enter into the peripheral circulation [24]. Proinflammatory cytokines, such as those implicated in this human ETEC challenge, were highly correlated in the peripheral blood and intestinal mucosa of ETEC-challenged pigs, further supporting the hypothesis that peripheral blood provides a useful correlate of tissue-specific changes [25]. In addition to ETEC infection, several studies revealed high correlation between mucosal and peripheral blood gene expression in human subjects with inflammatory bowel disease [26–28].

Drug repositioning identified candidates for further investigation, as well as novel mechanisms of pathogenesis. For example, the protein synthesis inhibitors anisomycin and puromycin induce gene expression changes in vitro similar to what is seen with ETEC infection, and both are produced by *Streptomyces* bacterial species. This suggests host protein synthesis inhibition is a part of ETEC pathogenesis as anisomycin does not have activity against *E. coli* [29]. Several antipsychotic drugs, which induce anti-inflammatory cytokine secretion and suppress proinflammatory cytokines [30], correlated with ETEC infection. This is consistent either with an ETEC-derived virulence mechanism or with attempts by the host to temper the inflammatory response. Tanespimycin, an antitumor antibiotic and Hsp90 inhibitor, was also identified as significant. ETEC increases TLR4 expression, whose function depends on Hsp90 [31]. Hsp90 inhibitors have also been investigated as therapeutics in intestinal parasitic diseases [32].

Symptomatic subjects demonstrated gene expression changes similar to those observed with NF-κB pathway inhibitors, a pathway associated with the ETEC response [33]. Another related pathway tied to ETEC pathogenesis is ubiquitination [34]. Withaferin A, significant in the repositioning analysis, is thought to inhibit NF-κB by inhibiting the ubiquitin-mediated proteasome pathway [17]. Although not classically associated with ETEC, type III secretion systems expressed by enteropathogenic and enterohemorrhagic *E. coli* inhibit host immune pathways, including NF-κB [35].

Zinc biology was prominent in the drug repositioning analysis. Vorinostat and trichostatin are histone deacetylase inhibitors that

chelate zinc [36]. Zinc enhanced innate immunity to ETEC in children, increasing complement C3 levels, enhancing phagocytic activity, and increasing the naive-memory T-cell ratio [37]. In a pig model, zinc supplementation decreased immune-response gene expression as well as expression of *MUC4,* a proposed ETEC K88 receptor [24].

Cephalexin is an antibiotic that can be used to treat ETEC infection. In the drug repositioning analysis, cephalexin had a negative enrichment score, suggesting possible symptom-mitigating effects on the host independent of direct antibacterial activity. Amoxicillin revealed similar associations. In contrast, rifampicin had a positive enrichment score, suggesting a potential immune augmentation benefit to the host, in addition to direct antibacterial activity. In light of the emergence of drug-resistant ETEC [38] and antibacterial resistance in general, therapeutic benefits achieved through host effects may become increasingly important.

Human host susceptibility to enteric infections has recently been described for norovirus and rotavirus, derived from polymorphisms in the *FUT2* gene, which represent the histo-blood group antigens to which both viruses bind [39]. Moreover, asymptomatic shedding of norovirus in outbreak situations has also been described, further supporting the importance of host factors in determining whether a given exposure results in disease [40]. In this study, we identified potential host susceptibility factors without a priori assumptions regarding their mechanisms. In doing so, we identified 29 genes differentially expressed at baseline that discriminated between symptomatic and asymptomatic individuals, suggesting innate resilience to ETEC infection. This finding offers insights into the biology of infection resilience, as well as potential targets for prognostic testing and therapeutic development. We acknowledge the potential for overfitting, making such discoveries potentially spurious. However, there is ample support for many of these genes playing a role in resilience. For example, *C4BPA* inhibits the classic complement pathway and is bound by a K1 *E. coli* strain outer membrane protein to evade complement-mediated destruction [41]. Its down-regulation in resilient subjects suggests an important role for complement activation in ETEC clearance. *GZMH,* up-regulated in resilient subjects, has antiviral proteolytic activity [42] and may also play a role in ETEC resilience.

Interestingly, Sutai pigs have variable ETEC infection susceptibility attributed in part to variable *E. coli* F18 adhesion [43], *TAP1* genotype [44], and IMTGP, an *E. coli* K88 receptor [45]. Although these specific genes and mechanisms were not identified in this study, we did note some overlap between porcine and human ETEC resistance mechanisms. For example, *CD8A,* part of the class I MHC pathway, was up-regulated in ETEC-resilient subjects as it was in resilient pigs. *SLA-1* and *SLA-3,* both involved in MHC antigen presentation, were also differentially expressed between sensitive and resistant pigs [21]. At the pathway level, immune and leukocyte activation were important to pig resilience. Although these pathways were not specifically enriched for in this human resilience signature, they were significant in the comparison of symptomatic versus asymptomatic subjects.

Some of the 29 resilience genes have also been identified in other infection-related gene expression studies. Similar to what we observed with ETEC, *LGALS8, TAP2,* and *WARS* were up-regulated in an influenza challenge experiment among subjects with more severe symptoms [46]. *WARS* was expressed at a higher level in patients with active versus latent tuberculosis [47]. In a *Haemophilus ducreyi* challenge study, *LGALS8, TAP2,* and *CD8A* were expressed at a higher level in subjects who became infected compared with those who did not, despite prior pathogen exposure [48]. We also found supportive evidence in an influenza vaccination model, hypothesizing that vaccination imparts resilience. *MXRA7* and *RNF187* showed similar expression changes in ETEC-susceptible subjects as in a prevaccination cohort (vs post-vaccination) [49]. Although these studies support the hypothesis that host resilience to infectious disease is detectable in the baseline healthy state, validation at a population scale or through functional experimentation would lend further support.

Limitations

This study included a small number of subjects, which limits generalizability. Nevertheless, the abundance of statistically significant and biologically plausible differences highlight the host response to ETEC challenge. Children bear the highest disease burden from ETEC but were not represented in this challenge. Although it cannot be assumed the pediatric and adult responses to ETEC are the same, their host responses are similar in other infectious diseases [3, 50]. Another caveat is that we measured peripheral blood gene expression and used this as a surrogate for the mucosal/intestinal host response. Although they are likely to be highly correlated, there are also likely to be differences. Lastly, the resilience signature was identified with the baseline measure obtained 8 hours after challenge. Although few changes are likely to be occurring this quickly based on other challenge experiments [3], the resilience gene expression signature may instead reflect an early, differential host response conferring resilience, rather than an innate biology.

Conclusions

This study is the first to describe statistically significant and biologically plausible differences in host gene expression induced by ETEC infection. Further, the differential expression of some genes at baseline may confer resilience to infection. These findings offer a baseline understanding of host response in gastrointestinal infection, which can serve as a basis for the study of other infectious diseases as well as noninfectious gastrointestinal disease.

Supplementary Data

Supplementary materials are available at http://jid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted

materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

Notes

Disclaimer. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Financial support. Funding for this study was provided by the US Defense Advanced Research Projects Agency (DARPA; contract N66001-07-C-2082). The volunteer study was supported by a grant from Program for Appropriate Technology.

Potential conflicts of interest. M. T. M. and E. L. T. were supported by the Clinical Science Research and Development Service of the VA Office of Research and Development (awards 1IK2CX000530 and 1IK2CX000611, respectively). G. S. G. has consulted for US Diagnostic Standards; has served on the scientific advisory board for Pappas Ventures; and has received grants from DARPA, the Gates Foundation, and Novartis Vaccines and Diagnostics. G. S. G, C. W. W., and E. L. T. have a patent pending for host gene expression signatures of *Staphylococcus aureus* and *E. coli* infections and have filed patents for methods of identifying infectious disease and assays for identifying infectious disease, as well as for molecular predictors of fungal infection. C. W. W. served as a scientific consultant to bioMerieux, Becton Dickinson, and Verigene and has received research support from the National Institutes of Health, DARPA, the Defense Threat Reduction Agency, the Bill and Melinda Gates Foundation, the Veterans Administration, the Centers for Disease Control and Prevention, Novartis Pharmaceuticals, Roche Molecular, bioMerieux, and Qiagen. E. L. T. has received research support from DARPA, the Defense Threat Reduction Agency, the Gates Foundation, the Veterans Administration, and Novartis Pharmaceuticals. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Porter CK, Riddle MS, Tribble DR, et al. A systematic review of experimental infections with enterotoxigenic *Escherichia coli* (ETEC). Vaccine **2011**; 29:5869–85.
2. Lindsay BR, Chakraborty S, Harro C, et al. Quantitative PCR and culture evaluation for enterotoxigenic *Escherichia coli* (ETEC) associated diarrhea in volunteers. FEMS Microbiol Lett **2014**; 352:25–31.
3. Zaas AK, Chen M, Varkey J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. Cell Host Microbe **2009**; 6:207–17.
4. Harro C, Chakraborty S, Feller A, et al. Refinement of a human challenge model for evaluation of enterotoxigenic *Escherichia coli* vaccines. Clin Vaccine Immunol **2011**; 18:1719–27.
5. West M. Bayesian factor regression models in the "large p, small n" paradigm. In: Bayesian statistics. Valencia, Spain. Oxford University Press, 723–32.
6. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, **2014**.
7. Bates D, Maechler M, Bolker B, Walker S. lme4: linear mixed-effects models using "Eigen" and S4. **2014**. https://cran.r-project.org/web/packages/lme4/citation.html. Accessed 2 January 2015.
8. Ritchie ME, Phipson B, Wu D, et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res **2015**; 43:e47.
9. Smedley D, Haider S, Durinck S, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res **2015**; 43:W589–98.
10. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **2009**; 4:44–57.
11. Levine MM, Nalin DR, Hoover DL, Bergquist EJ, Hornick RB, Young CR. Immunity to enterotoxigenic *Escherichia coli*. Infect Immun **1979**; 23:729–36.
12. Rao MR, Wierzba TF, Savarino SJ, et al. Serologic correlates of protection against enterotoxigenic *Escherichia coli* diarrhea. J Infect Dis **2005**; 191:562–70.
13. Safran M, Dalah I, Alexander J, et al. GeneCards version 3: the human gene integrator. Database **2010**; 2010:baq020.
14. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res **2014**; 42:D553–9.
15. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science **2006**; 313:1929–35.
16. Straus DS, Pascual G, Li M, et al. 15-deoxy-$\Delta^{12,14}$-prostaglandin J_2 inhibits multiple steps in the NF-κB signaling pathway. Proc Natl Acad Sci U S A **2000**; 97:4844–9.
17. Mohan R, Hammers HJ, Bargagna-Mohan P, et al. Withaferin A is a potent inhibitor of angiogenesis. Angiogenesis **2004**; 7:115–22.
18. Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med **2015**; 7:287ra71.
19. Ramilo O, Allman W, Chung W, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood **2007**; 109:2066–77.
20. Jain U, Cao Q, Thomas NA, et al. Properdin provides protection from *Citrobacter rodentium*-induced intestinal inflammation in a C5a/IL-6-dependent manner. J Immunol **2015**; 194:3414–21.
21. Ye L, Zi C, Pan ZY, et al. Investigation of the relationship between SLA-1 and SLA-3 gene expression and susceptibility to *Escherichia coli* F18 in post-weaning pigs. Comp Immunol Microbiol Infect Dis **2012**; 35:23–30.
22. Bao WB, Ye L, Pan ZY, et al. Microarray analysis of differential gene expression in sensitive and resistant pig to *Escherichia coli* F18. Anim Genet **2012**; 43:525–34.
23. Zhu YH, Li XQ, Zhang W, Zhou D, Liu HY, Wang JF. Dose-dependent effects of *Lactobacillus rhamnosus* on serum interleukin-17 production and intestinal T-cell responses in pigs challenged with *Escherichia coli*. Appl Environ Microbiol **2014**; 80:1787–98.
24. Sargeant HR, McDowall KJ, Miller HM, Shaw MA. Dietary zinc oxide affects the expression of genes associated with inflammation: transcriptome analysis in piglets challenged with ETEC K88. Vet Immunol Immunopathol **2010**; 137:120–9.
25. Nyachoti CM, Kiarie E, Bhandari SK, Zhang G, Krause DO. Weaned pig responses to *Escherichia coli* K88 oral challenge when receiving a lysozyme supplement. J Anim Sci **2012**; 90:252–60.
26. Song L, Zhou R, Huang S, et al. High intestinal and systemic levels of interleukin-23/T-helper 17 pathway in Chinese patients with inflammatory bowel disease. Mediators Inflamm **2013**; doi:10.1155/2013/425915.
27. Kim SW, Kim ES, Moon CM, Kim TI, Kim WH, Cheon JH. Abnormal genetic and epigenetic changes in signal transducer and activator of transcription 4 in the pathogenesis of inflammatory bowel diseases. Dig Dis Sci **2012**; 57:2600–7.
28. Shen J, Qiao YQ, Ran ZH, Wang TR. Up-regulation and pre-activation of TRAF3 and TRAF5 in inflammatory bowel disease. Int J Med Sci **2013**; 10:156–63.
29. Grollman AP. Inhibitors of protein biosynthesis. II. Mode of action of anisomycin. J Biol Chem **1967**; 242:3226–33.
30. Al-Amin MM, Nasir Uddin MM, Mahmud Reza H. Effects of antipsychotics on the inflammatory response system of patients with schizophrenia in peripheral blood mononuclear cell cultures. Clin Psychopharmacol Neurosci **2013**; 11:144–51.
31. Finamore A, Roselli M, Imbinto A, Seeboth J, Oswald IP, Mengheri E. Lactobacillus amylovorus inhibits the TLR4 inflammatory signaling triggered by enterotoxigenic *Escherichia coli* via modulation of the negative regulators and involvement of TLR2 in intestinal Caco-2 cells and pig explants. PloS One **2014**; 9:e94891.
32. Debnath A, Shahinas D, Bryant C, et al. Hsp90 inhibitors as new leads to target parasitic diarrheal diseases. Antimicrob Agents Chemother **2014**; 58:4138–44.
33. Wang X, Gao X, Hardwidge PR. Heat-labile enterotoxin-induced activation of NF-κB and MAPK pathways in intestinal epithelial cells impacts enterotoxigenic *Escherichia coli* (ETEC) adherence. Cell Microbiol **2012**; 14:1231–41.
34. Wang X, Hardwidge PR. Enterotoxigenic *Escherichia coli* prevents host NF-κB activation by targeting IκBα polyubiquitination. Infect Immun **2012**; 80:4417–25.
35. Coburn B, Sekirov I, Finlay BB. Type III secretion systems and disease. Clin Microbiol Rev **2007**; 20:535–49.
36. Richon VM. Cancer biology: mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor. Br J Cancer **2006**; 95:S2-6.
37. Sheikh A, Shamsuzzaman S, Ahmad SM, et al. Zinc influences innate immune responses in children with enterotoxigenic *Escherichia coli*-induced diarrhea. J Nutr **2010**; 140:1049–56.
38. Cruz-Cordova A, Espinosa-Mazariego K, Ochoa SA, et al. CS21 positive multidrug-resistant ETEC clinical isolates from children with diarrhea are associated with self-aggregation, and adherence. Front Microbiol **2014**; 5:709.
39. Kambhampati A, Payne DC, Costantini V, Lopman BA. Host genetic susceptibility to enteric viruses: a systematic review and metaanalysis. Clin Infect Dis **2016**; 62:11–8.
40. Costantini VP, Cooper EM, Hardaker HL, et al. Epidemiologic, virologic, and host genetic factors of norovirus outbreaks in long-term care facilities. Clin Infect Dis **2015**; 62:1–10.

41. Wooster DG, Maruvada R, Blom AM, Prasadarao NV. Logarithmic phase *Escherichia coli* K1 efficiently avoids serum killing by promoting C4bp-mediated C3b and C4b degradation. Immunology **2006**; 117:482–93.
42. Waterhouse NJ, Trapani JA. H is for helper: granzyme H helps granzyme B kill adenovirus-infected cells. Trends Immunol **2007**; 28:373–5.
43. Meijerink E, Fries R, Vogeli P, et al. Two α(1,2) fucosyltransferase genes on porcine chromosome 6q11 are closely linked to the blood group inhibitor (*S*) and *Escherichia coli* F18 receptor (*ECF18R*) loci. Mamm Genome **1997**; 8:736–41.
44. Zhao Q, Liu Y, Dong W, et al. Genetic variations of *TAP1* gene exon 3 affects gene expression and *Escherichia coli* F18 resistance in piglets. Int J Mol Sci **2014**; 15:11161–71.
45. Francis DH, Grange PA, Zeman DH, Baker DR, Sun R, Erickson AK. Expression of mucin-type glycoprotein K88 receptors strongly correlates with piglet susceptibility to $K88^+$ enterotoxigenic *Escherichia coli,* but adhesion of this bacterium to brush borders does not. Infect Immun **1998**; 66:4050–5.
46. Davenport EE, Antrobus RD, Lillie PJ, Gilbert S, Knight JC. Transcriptomic profiling facilitates classification of response to influenza challenge. J Mol Med **2015**; 93:105–14.
47. Maertzdorf J, Repsilber D, Parida SK, et al. Human gene expression profiles of susceptibility and resistance in tuberculosis. Genes Immun **2011**; 12:15–22.
48. Humphreys TL, Li L, Li X, et al. Dysregulated immune profiles for skin and dendritic cells are associated with increased host susceptibility to *Haemophilus ducreyi* infection in human volunteers. Infect Immun **2007**; 75:5686–97.
49. Parnell G, McLean A, Booth D, Huang S, Nalos M, Tang B. Aberrant cell cycle and apoptotic changes characterise severe influenza A infection—a meta-analysis of genomic signatures in circulating leukocytes. PloS One **2011**; 6: e17186.
50. Mejias A, Dimo B, Suarez NM, et al. Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. PLoS Med **2013**; 10:e1001549.